AR-based video-mediated communication:

A social presence enhancing experience

Igor de Souza Almeida,
Marina Atsumi Oikawa, Jordi Polo Carres,
Jun Miyazaki, Hirokazu Kato
Nara Institute of Science and Technology
Nara, Japan
{igor-a, marina-o, jordi-p, miyazaki, kato}@is.naist.jp

Mark Billinghurst
University of Canterbury
Canterbury, New Zealand
mark.billinghurst@hitlabnz.org

***Abstract*— Video-mediated communication systems attempt to provide users with a channel that could bring out the "feeling" of face-to-face communication. Among the many qualities these systems aim for, a high level of Social Presence is unquestionably a desirable one; however, little effort has been made to improve upon the user's perception of "presence". We propose an AR approach to enhance social presence for video-mediated systems by allowing one user to be present in the other user's video image. We conducted a preliminary pilot study with 10 participants coupled in 5 pairs to evaluate our system and compare with the traditional video-chat setup. Results indicated that our system has higher degree of social presence compared to traditional video-chat systems. This conclusion was supported by the positive feedback from the subjects.**

Keywords- augmented reality; social presence; video-mediated communication.

I. INTRODUCTION

Video mediated communication systems are widely spread and easily accessible in their "video-chat" form which is the reason why it is one of the most popular communication channels being used between remotely located people.

Traditional video-chat systems rely primarily on the two-way video/audio feedback in order to cover the cues for an efficient human-human communication. However, taking into account that face-to-face communication is the optimal communication case, there are a number of natural non-verbal communication cues which help one to convey the message, including gaze direction, proximity behavior and pointing in space [1][2].

For example, during Mother's Day, a user is video chatting with his/her mother. Even though they can listen and see each other, the user may wish that he/she could touch her or give her a gift, just like in a face-to-face meeting. Researchers have since explored new concepts to offer a more complete experience. The idea of Shared Space is one of them. Shared space draws similarities with Collaborative Virtual Environments (CVE) in which participants and information share a common display space [3]. Even though it has been proven that using a dedicated shared space environment improves the results of certain tasks [4][5], it does not address the effect of having an additional separated environment aside from the users' real world.

In HyperMirror [6], a shared scene is created by capturing the front view of users in different rooms and merging the images of all users in one video image. The resulting image displays all users as if they were side-by-side in front of a mirror; users have to keep facing forward at all times making it difficult to interact with another user that is virtually by their side.

In this work, we intend to improve upon the traditional video-chat setup which typically consists of two windows being presented to both ends: one to display the user's own video image and another one to show the remote person. By adding one extra camera focusing on a shared space with monochromatic background, a user can place his/her hand/s inside this space and have them combined with the other user's face image. This simulates their coexistence and stimulates different interaction patterns as previously reported by [7]. We advocate that this has the potential to enhance social presence.

Social presence has been defined as the salience of the partner in a mediated communication and the consequent salience of their interpersonal interactions [8]. However, [9] also distinctively points out that social presence has been shown to relate more to the user's perception of a medium's ability to provide salience of another as opposed to measuring the actual perceived salience of another person.

In this work, social presence is targeted as the measure to perceive how much the presence of one user's video image into another user's video image can enhance sense of "being part of" each other's environment. Similar work can be found in [10][11]. The latter compares a video conferencing system with an Augmented Reality chat in terms of Social Presence and Co-presence. No significant differences were found for neither aspects.

II. HANDY SYSTEM

In our proposed system, the main idea is to minimize the feeling of being geographically separated by allowing one user to be part of the video image of the other user.

Figure 1 shows a comparison between the setup of a traditional video-chat system and our proposed system, referred to as HANDY. In the former, User B (in red background) has one camera capturing his face image which

978-0-7695-4725-1/25 $26.00 © 4725 IEEE
DOI 10.1109/SVR.2012.4

is displayed in the small window on top of User A's face image (in blue background) in the big window. Users are restricted to interact in their own environment.

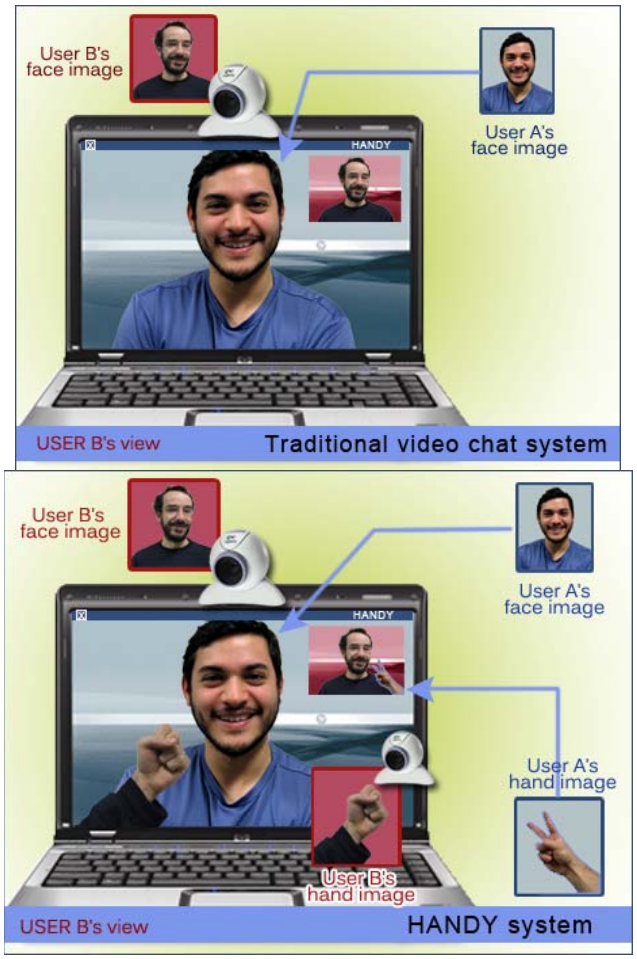

Figure 1. Conceptual images of the setup for a traditional video chat system (above) and HANDY system (below) considering user B's view.

In the latter, HANDY system, there is an extra camera focusing on the hand gestures alongside the camera in the traditional setup. The hand image is then sent and merged into the other user's face image. In Figure 1, it is possible to see the hand of User A merged with User B's face image in the small window and the hand of User B merged with User A's face image in the big window, as if they had crossed to the other side.

Figure 2 shows the real environment setup for HANDY. Basically, each user has two cameras in the environment: one placed towards the face, in the same position used in conventional video-chat system; another placed beside the user, capturing the area where the hands can be inserted. This area consists of a monochrome static background and is referred to as shared space. Even though users never actually share this space, this is the area that allows them to share each other's real environments.

The shared space requires a training phase using the second camera (hand camera) using the first n initial frames containing the monochrome background. The training step is necessary to allow the segmentation of foreground and background pixels from an incoming frame according to a threshold value t (Figure 3).

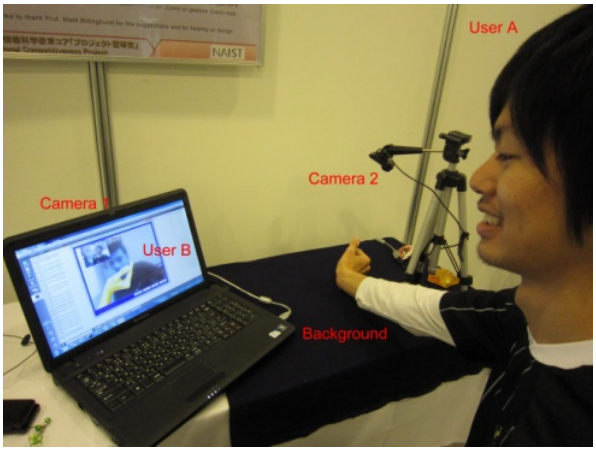

Figure 2. Camera 1 captures the face image and Camera 2 captures the hand image of User A. While using HANDY, the face image of User B is showed in the computer's screen merged with the User A's hand image.

After this segmentation, the background is subtracted and the image's parts of interest (the hands) are merged with the remote user's video image in real time. OpenCV library was chosen to deal with the capture and merging of local and remote video images and Qt library was used to develop the entire communication architecture for the dual video streams.

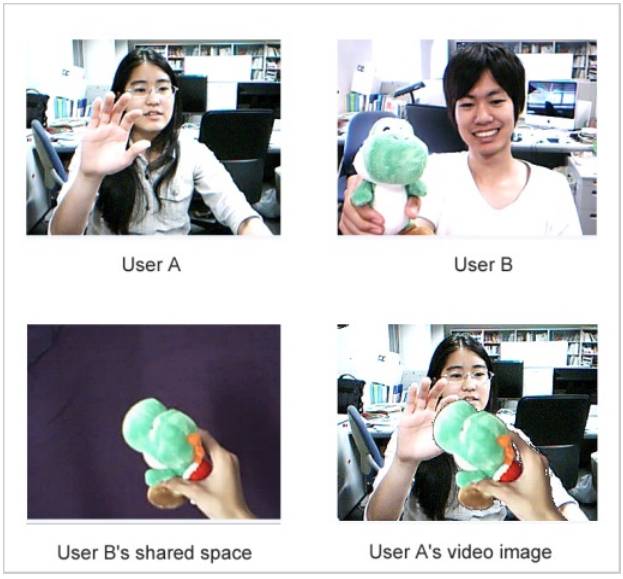

Figure 3. Background subtraction technique

The use of a simple technical solution made it affordable when applying it to video-chat system in contrast to more complex solutions such as pure Augmented Reality. In addition, it accelerated the prototyping process. The idea is to demonstrate the potential of the system by achieving equivalent visual results without the need to implement all

the complexity involved in the pure AR solution. Nevertheless, the use of a high quality virtual image along with the proper 3D registration could potentially bring a more polished and highly interactive result.

III. User Interaction

In this section, we describe a practical scenario for the use of HANDY in comparison to the traditional video-chat system. The view shown in Figure 4 and Figure 5 is from User A's side.

A. Scenario: giving an object

This scenario corresponds to a very simple action in which User A would like to give an object to User B. In HANDY (Figure 5), user A would initially show the object to user B as it would happen in the traditional system. However, as soon as user A places the object in the shared space, it would appear on user B's environment. The effect of seeing an object that was on user A's video image being then part of user B's video image is one of the motivational factors that could impact social presence [12].

Figure 4. Giving a gift in a traditional video-chat setting.

Figure 5. Giving a gift using HANDY.

IV. Evaluation

We designed a pilot user study using HANDY against the traditional video-chat system to evaluate the degree of Social Presence as well as Ease of Use, Enjoyment, Ease of Communication, Intuitiveness, Ease of understanding and Closeness. The study's scenario consisted mainly of remote users performing interactive tasks under equal roles, being the exception the final task where subjects played different roles (teacher/student).

A. Participants

We had a total of 10 participants in the study comprising 8 Japanese students and 2 foreign students, 23 to 35 years old (average 25.5). They were all male graduate students from the Interactive Media Design Lab and have experienced Augmented Reality before.

All participants had met before the experiment which helped creating a relaxed atmosphere. Furthermore, since they were same-sex pairs, any tensions that may arise from inter-gender interactions were disregarded. The participants performed in pairs over two days with each session taking about one hour.

For each pair, each participant was taken to a separate room and seated in front of a large display. Initially, they were asked to fill out a demographic survey which was followed by a brief walkthrough of the study by an experiment assistant. Participants were instructed on how to use HANDY.

A questionnaire was given after each condition for every task. Another questionnaire was given after both conditions had been performed for them to evaluate comparatively both conditions under that given task. A final survey was given at the end of the session for the participants to evaluate comparatively both conditions considering the experience as a whole.

B. Tasks

Three tasks were performed by each pair. The first task was a Rock-paper-scissors game. Participants were asked to play six games for each condition. Even though the wins were counted, a winner was not declared at the end.

The second task was a Puzzle Matching (Figure 6a) where participants were given an equal number of puzzle pieces (six). Each piece had a unique match (unique connecting edges), once assembled they all completed the same image, a rabbit. Their task was to find the match for each puzzle. One subject was given six pieces containing the rabbit's head and the other subject was given six pieces containing the rabbit's body. They used the same pieces for both conditions.

The third one was an Origami Training task (Figure 6b). One subject was given the role of a Teacher while the other played the role of a Student. The Teacher was given a print out containing the instructions on how to fold an origami. Two different origamis with distinct degrees of difficulty were chosen: a piano (easy level) and a balloon (medium level). For each pair, the order of the selected origami for each condition was changed. For this task, the cameras were targeting the hands on the table from a top view instead of

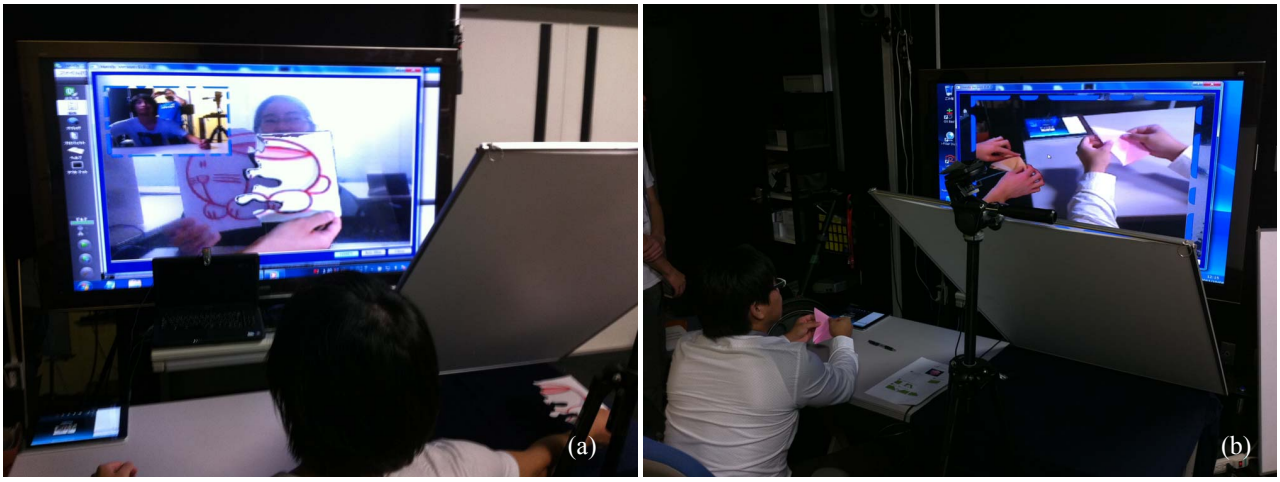

Figure 6. Pictures taken during the experiments: (a) Puzzle Matching (b) Origami Training.

the face. In addition, the only video image being displayed was the one with the merging.

C. Conditions

- **HANDY Off**: this condition simulated the traditional video-chat setup where real-time audio and video feed are available to the subjects only.
- **HANDY On**: this condition assumed the same setup as HANDY Off, but also adding the merging of one subject's hand image (captured by a second webcam placed on his/her side and facing a monochrome background) with the other subject's video image ("face image").

During condition HANDY On, subjects were asked to use HANDY alternately to have both of them experiencing the system equally. For Task 3, however, only the user who took the Student role used HANDY actively, placing the hand in front of the second webcam.

D. Experimental design

We used a within-subjects design with a single independent variable (HANDY On & HANDY Off). The order of the tasks was the same throughout the experiments. Observation notes were taken while watching participants' performance in each task to record different interaction and behavior patterns.

In this experiment we used semantic differential measure for measuring Social Presence with focus on the medium. This was described by Short et al. [8], referred by [13] as "the most commonly used measure of Social Presence" and used by [1][11]. We used nine bipolar scales including items such as impersonal/personal, cold/warm among others.

Likert Scales (7-point) were used in the exploratory questionnaire items to evaluate the overall performance of the system according to six measures (Ease of Use, Enjoyment, Ease of Communication, Intuitiveness, Ease of understanding and Closeness).

Our hypotheses about the study's outcome were mainly targeting Social Presence and they can be described as follows:

$H1_0$: HANDY On and HANDY Off generate a similar degree of measured social presence.

$H1_a$: There is a significant difference in the measurement of social presence created by the two conditions.

V. Results

A. On HANDY

We are looking primarily for indications that the prototype system could provide a satisfying communication platform as well as the users' feedback on the experience with the system.

Following completion of each task users filled out a subjective survey asking them to respond to Likert Scale questions on a range from 1 to 7, 1 = strongly disagree, 7 = strongly agree. A two factor (task, Handy On/Off) ANOVA test was performed on the results.

We found significant differences between conditions for Enjoyment ($F(2,54) = 4.019$, $P < 0.05$), Ease of Communication ($F(2,54) = 4.019$, $P < 0.05$) and Intuitiveness ($F(2,54) = 4.019$, $P < 0.05$). The average scores for these measures are shown in Figure 7. They show that there is a difference in user scores between the HANDY Off and HANDY On conditions.

For the puzzle matching and origami training tasks, users thought the HANDY On condition was more fun as well as it made the communication easier to be performed and more natural. These are aspects that could signal a possible higher degree of Social Presence under the given condition. For the rock-paper-scissor task however even though users thought the HANDY On condition was more fun, it was not more natural or easier to communicate with than HANDY Off condition.

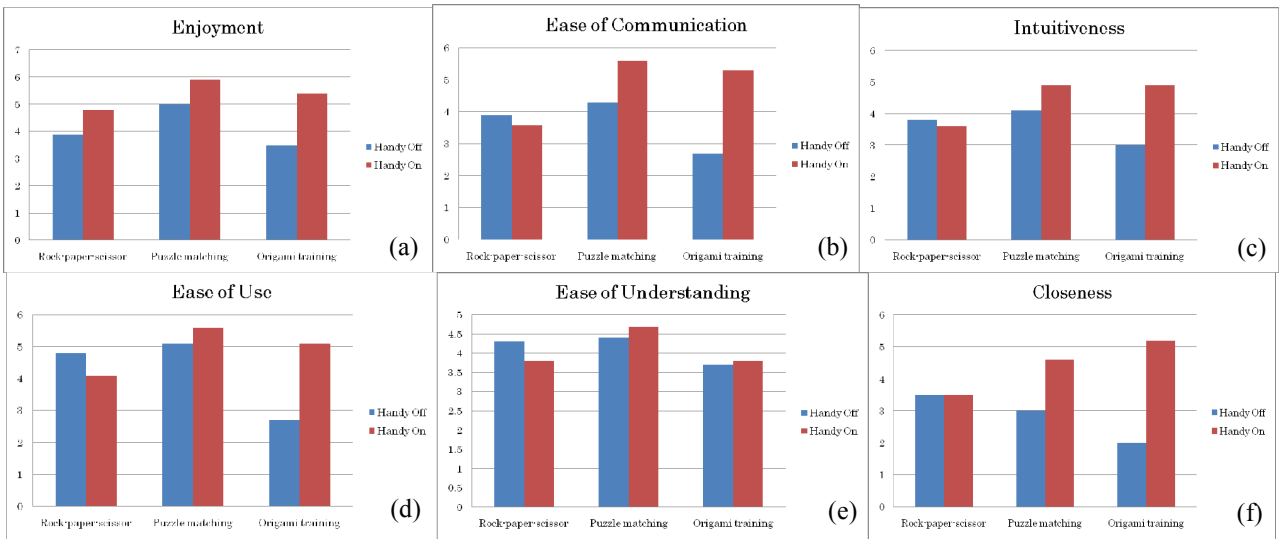

Figure 7. Average scores for the significant Likert scale measures: (a) Enjoyment, (b) Ease of Communication, (c) Intuitiveness, (d) Ease of use, (e) Ease of Understanding and (f) Closeness.

In addition to conducting a user survey we also asked users for their comments about the system. The comments reflected the results obtained from the questionnaires: most of the participants agree that when using HANDY system (HANDY On condition), the experience becomes more fun and it improves communication abilities.

Other comments however referred to non-ideal settings such as the position of the second camera. Ideally, this second camera should be placed behind the user's back and in a height close to the user's viewpoint. However, due to limitations when using the background subtraction (training stage), the camera was placed beside the user. Eye contact and some video latency were also reported as technical points that could affect the communication if the task relies on any of those factors. For example, the rock-paper-scissor task became slightly troublesome since the users had to adapt the timing of their hands due to latency.

B. On Social Presence

Once we validate the prototype and its capability to handle and deliver video-mediated communication, we started looking deeper into how much the users perceived each other through the system.

Following completion of each condition, users filled out subjective surveys asking them to respond to Bipolar Scale questions on a range from 1 to 7 (1 = negative end; 7 = positive end) evaluating nine factors regarding their feeling towards the medium: Impersonal/Personal, Cold/Warm, Ugly/Beautiful, Small/Large, Insensitive/Sensitive, Colourless/Colourful, Unsociable/Sociable, Closed/Open and Passive/Active. A two factor (task, Handy On/Off) ANOVA test was performed on the results.

Despite the small number of samples, significant differences were found between conditions HANDY On (Mean=4.5481, StDev=1.2707, $p<0.05$) and HANDY Off (Mean=3.5925, StDev=1.1423, $p<0.05$) across all nine factors. Figure 8 shows the means of social presence for each condition over the three tasks. The average of social presence was consistently rated higher during condition HANDY On.

Post-hoc comparisons showed that social presence was significantly higher in HANDY On condition, making the current results enough to at least indicate HANDY On's tendency towards having a higher degree of social presence by statistically rejecting the null hypothesis.

Further investigation is to target additional factors that could influence the social presence degree of HANDY such as gestures. In particular, the different types of gestures the user is successfully making use of for communication and/or tasks purposes as well as to what degree the system enables those gestures to be correctly expressed.

VI. DISCUSSION

Subjects had mixed feelings towards Task 1 (Rock-paper-scissors) and Task 2 (Puzzle Matching) tasks in contrast with the generally positive feedback given for the experience in Task 3 (Origami training). This may be due to the nature of the interaction involved in the given tasks.

Task 1 had a predetermined set of gestures which represent the options in the game. The hands were held in the air while showing the hand's shape. During the experiment, subjects were often trying to reposition their hands due to hands occlusion. Different from the face-to-face game, there is only one angle of view; therefore users need to rearrange their spaces to play the game successfully. Trying to find the right position in the 2D displayed image while moving the hand in the 3D real world was not an easy task.

Task 2 generated more active and open gestures. Different strategies arisen in order to perform the match more efficiently such as dropping the piece inside the shared space to make it static and describing the contour of the piece to the other user since trying to actively match the pieces brought up difficulties.

Task 3 included a sheet with instructions on how to fold the origami but it was up to the subject to decide how to pass that information without showing them explicitly. In preparation for the task, the users naturally adjusted their hands' position. In contrast to Task 1 and 2, it was like preparing your space on a table, similar to a face-to-face situation. In this case, occlusion was not frequent.

The complete focus on the hands for the entire time of Task 3 might have helped in the participants' higher involvement in the task, but it also provided less communication cues since no face image was being displayed.

The higher degree of social presence reported in all tasks indicates that the users perceived their partners more efficiently however it does not guarantee one feels he/she is "being perceived" simultaneously. In this case, a co-presence study would be needed in order to clarify this possibility.

The simultaneous use of HANDY by both users has not being investigated due to the fact that if HANDY is used simultaneously, the merging would be happening in both windows (of both users) causing loss of information which could possibly affect the communication.

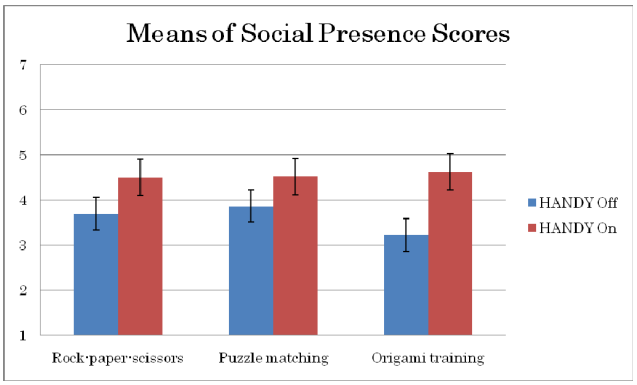

Figure 8. Average ratings of measured social presence.

Since this work is a prototype of a future AR application, it was important that the participants could perceive the technology. By having their real world being enhanced by virtual contents (partner's hands), the system achieved its goal. In spite of video latency and low resolution of the video images, the medium's ability to support the communication was seen as satisfactory.

Aside from modifications on the experimental design, a more reasonable amount of samples should be collected for statistically reliable analysis.

VII. CONCLUSION

We demonstrated an AR approach for enhancing social presence in video-chat communication and introduced a prototype system that allows one user into another user's video image. We presented both quantitative and qualitative results of a user study with 10 participants to evaluate the degree of social presence and exploratory factors. The proposed system has significant difference when compared to the traditional video-chat condition, including a higher social presence degree and more engaging experience. We conclude that our approach is promising and highlights the potential of AR technology in the traditional video-mediated communication context.

REFERENCES

[1] Joerg Hauber. Understanding Remote Collaboration in Video Collaborative Virtual Environments. Master's Thesis, University of Canterbury, Christchurch, New Zealand, 2008.

[2] Carmen Edigo. Teleconferencing as a technology to support cooperative work: Its possibilities and limitations. In Intellectual Teamwork: Social and Technological Foundations of Cooperative Work. Hillsdale, NJ: Lawrence Erlbaum Associates, 1990.

[3] Steve Benford et al. Shared Spaces: Transportation, Artificiality, and Spatiality. In Proceedings of CSCW '96, pages 77-86, 1996.

[4] Judith S. Olson, J.S, Gary M. Olson and David K. Meader. What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work? In Proceedings of the SIGCHI '95, pages 362-368, 1995.

[5] Steve Whittaker et al. Geelhoed, E. and Robinson E. Shared workspaces: how do they work and when are they useful? In International Journal of Man-Machine Studies, volume 39, pages 813-842. Academic Press Ltd, 1993.

[6] Osamu Morikawa and Takanori Maesako. HyperMirror: toward pleasant-to-use video mediated communication system. In Proceedings of CSCW '98, pages 149-158. ACM Press, 1998.

[7] Charlotte N. Gunawardena. Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences. In International Journal of Educational Telecommunications, volume 1, pages 147-166. Charlottesville, VA, AACE, 1995.

[8] John Short, Ederyn Williams and Bruce Christie. The social psychology of telecommunications. Wiley Ltd, London, UK, 1976.

[9] Eva-Lotta Sallnas. Effects of Communication Mode on Social Presence, Virtual Presence and Performance in Collaborative Virtual Environments. In Presence: Teleoperators and Virtual Environments, volume 14, pages 434-449. MIT Press, Cambridge, USA, 2005.

[10] Erin Bradner and Gloria Mark. Social presence with video and application sharing. In Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work, pp. 154-161. Boulder, Colorado, USA: ACM Press, 2001.

[11] Travis B. Faas and James Mohler. An Examination of Social Presence in Video Conferencing vs. an Augmented Reality Conferencing Application. Master's Thesis, Purdue University, Indiana, United States of America, 2010.

[12] Claus J. Knudsen. Video Mediated Communication (VMC) - Producing a sense of presence between individuals in a shared virtual reality. In Proceedings of International Symposium on Educational Conferencing. 2002.

[13] Frank Biocca, Chad Harms and Judee K. Burgoon. Toward a more robust theory and measure of social presence: review and suggested criteria. In Presence: Teleoperators and Virtual Environments, volume 12, pages 456-480. 2003.